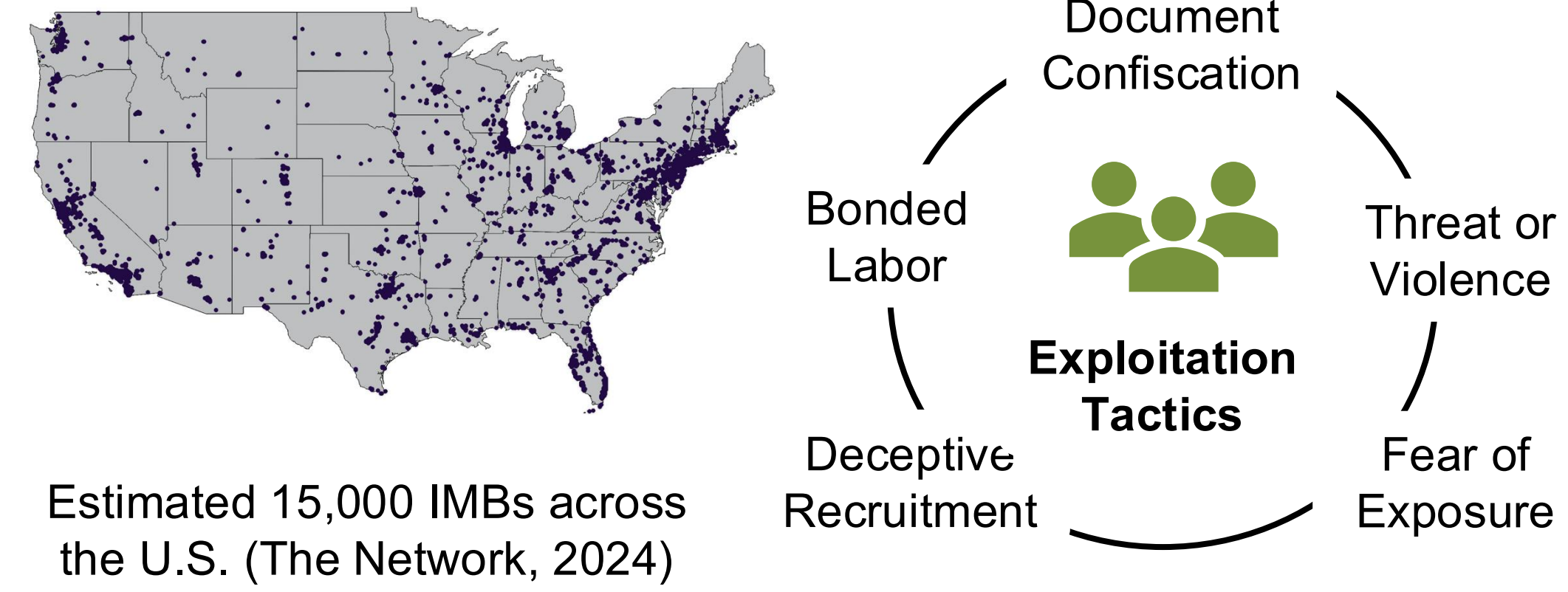
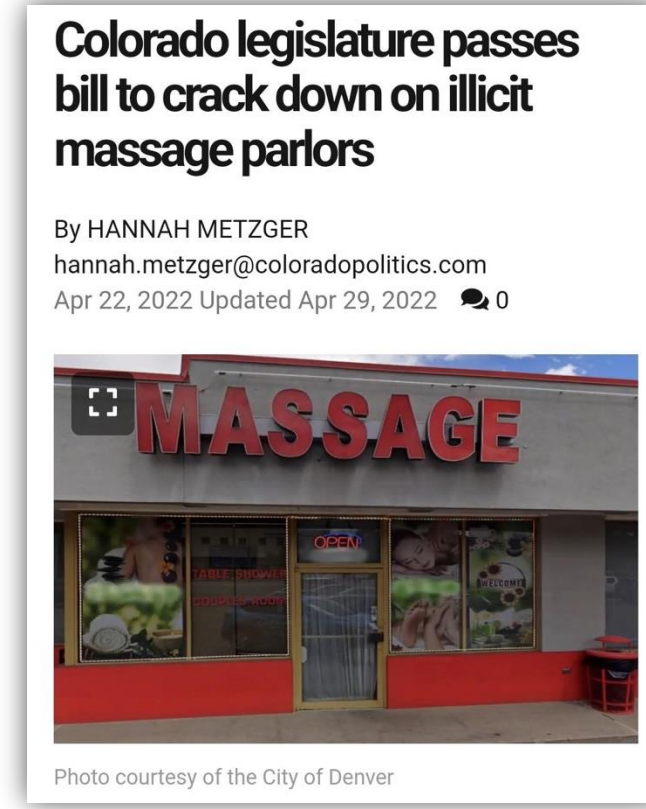
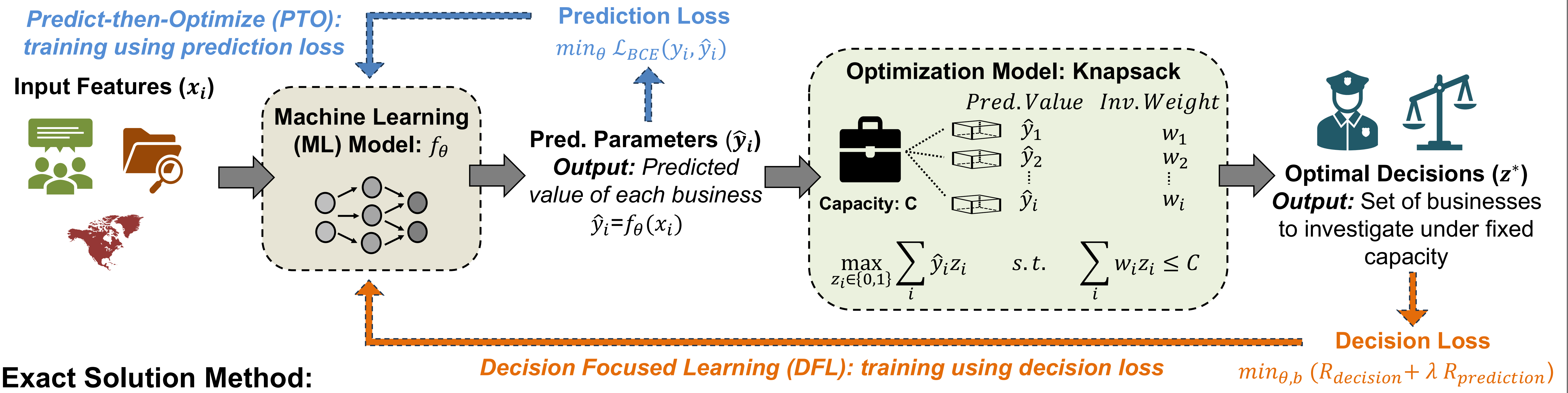


Introduction

- Illicit massage businesses (IMBs) across the U.S. offer commercial sex**
- These illegal establishments exploit “workers” through **sex and labor trafficking**
- Law Enforcement Agencies (LEAs) have limited resources** and are overwhelmed by the **volume** of investigations
- We integrate **predictive analytics (ML)** with **prescriptive decision-making** for scalable, **detection** of IMBs with the **highest impact** under limited capacity



Methodology: Decision Focused Learning



Exact Solution Method:

Objective:

$$\min_{\theta, b, z_{ki}} \sum_{k=1}^K \sum_{i=1}^n (y_{ki} z_{ki}^* - y_{ki} z_{ki} + \lambda(1 - y_{ki}) \hat{y}_{ki}^{bin} + \lambda y_{ki} (1 - \hat{y}_{ki}^{bin}))$$

$R_{decision}$: Lost Utility $R_{prediction}$: FP + FN

Indices: $k = 1, \dots, K$ (knapsack instance), $i = 1, \dots, n$ (business), $j = 1, \dots, p$ (feature)

Key Constraints:

- Machine Learning Predictions: $\hat{y}_{ki} = b + \sum_j \theta_j x_{kij}$
- Binary Thresholding $\hat{y}_{ki}^{bin} = 1$, where $\hat{y}_{ki} \geq (thresh)$; 0 o.w.
- Investigative Weight Ordering: $z_{ki1} \geq z_{ki2} - (1 - \hat{y}_{ki1}^{bin}) - (1 - \hat{y}_{ki2}^{bin})$
- Selection Constraint: $z_{ki} \leq \hat{y}_{ki}^{bin}$
- Investigative Capacity: $\sum_i w_{ki} z_{ki} \leq C$

Multi-source Integration - Data

Labels:

- RubMap.ch & License Disciplinary data

Business (focused on Colorado):

- Yelp, GIS, U.S. Census, Land Cover type
- Total: 425 (Label 1: 85, Label 0: 340)

Investigative Weights:

- Licensing, out-of-state links
- Hotspot proximity, demographics

Knapsack Instantiation:

- Randomly sampled 50 businesses for 100 knapsack instances

Deployment: Testing Results

To get optimal decisions (z_{ki}) from trained ML parameters (\hat{y}_{ki}):
Solving knapsack with Objective: $\max_{z_{ki}} \sum_{k=1}^K \sum_{i=1}^n \hat{y}_{ki} z_{ki}$, **Constraint:** $\sum_i w_{ki} z_{ki} \leq C$

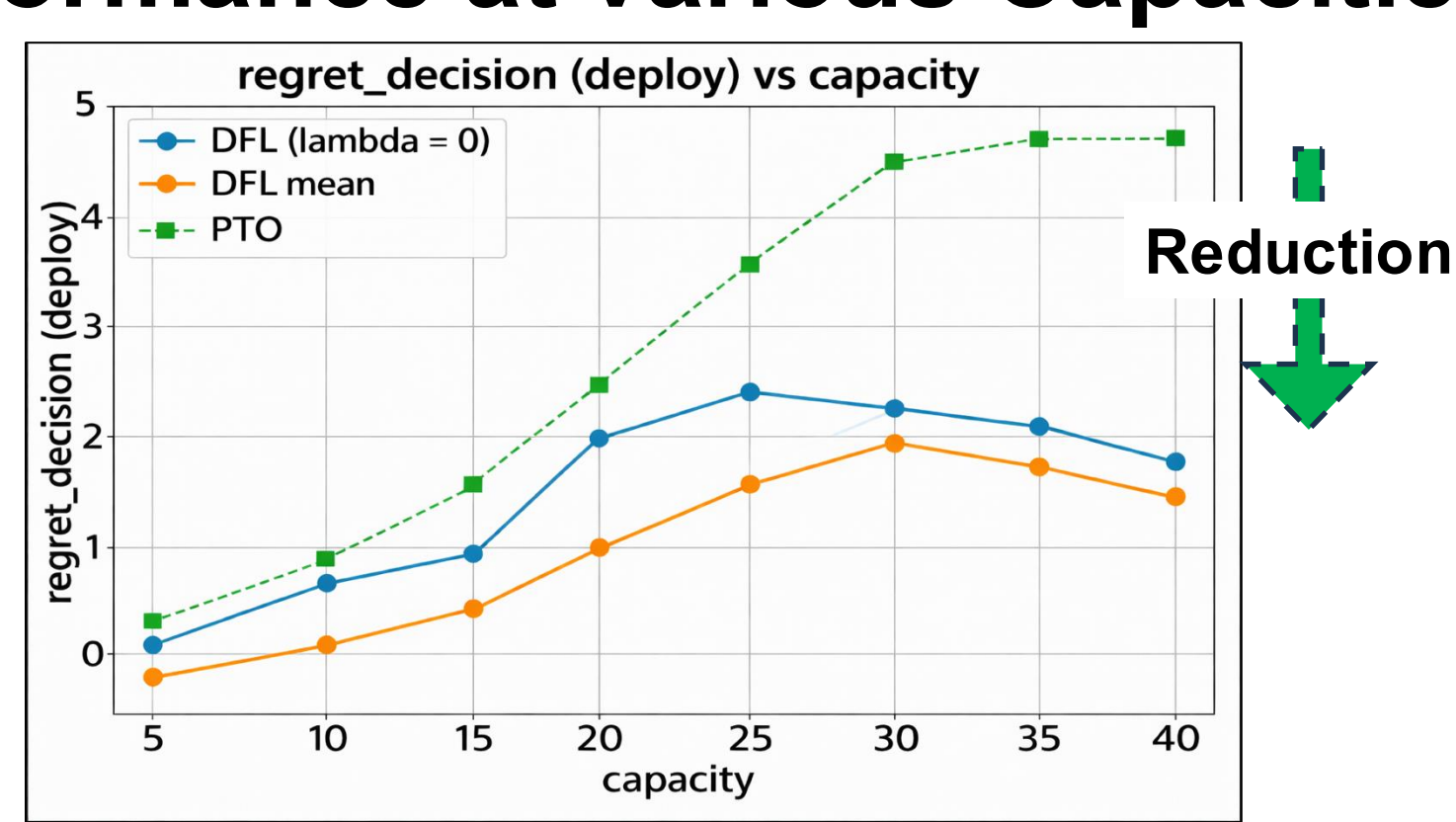
Performance Metrics

C	λ	Method	$R_{decision}$	$R_{prediction}$
15	0	PTO	2.0750	9.6000
15	0	DFL	1.2250	10.2750
15	0.5	PTO	2.0750	9.6000
15	0.5	DFL	0.7750	3.3250

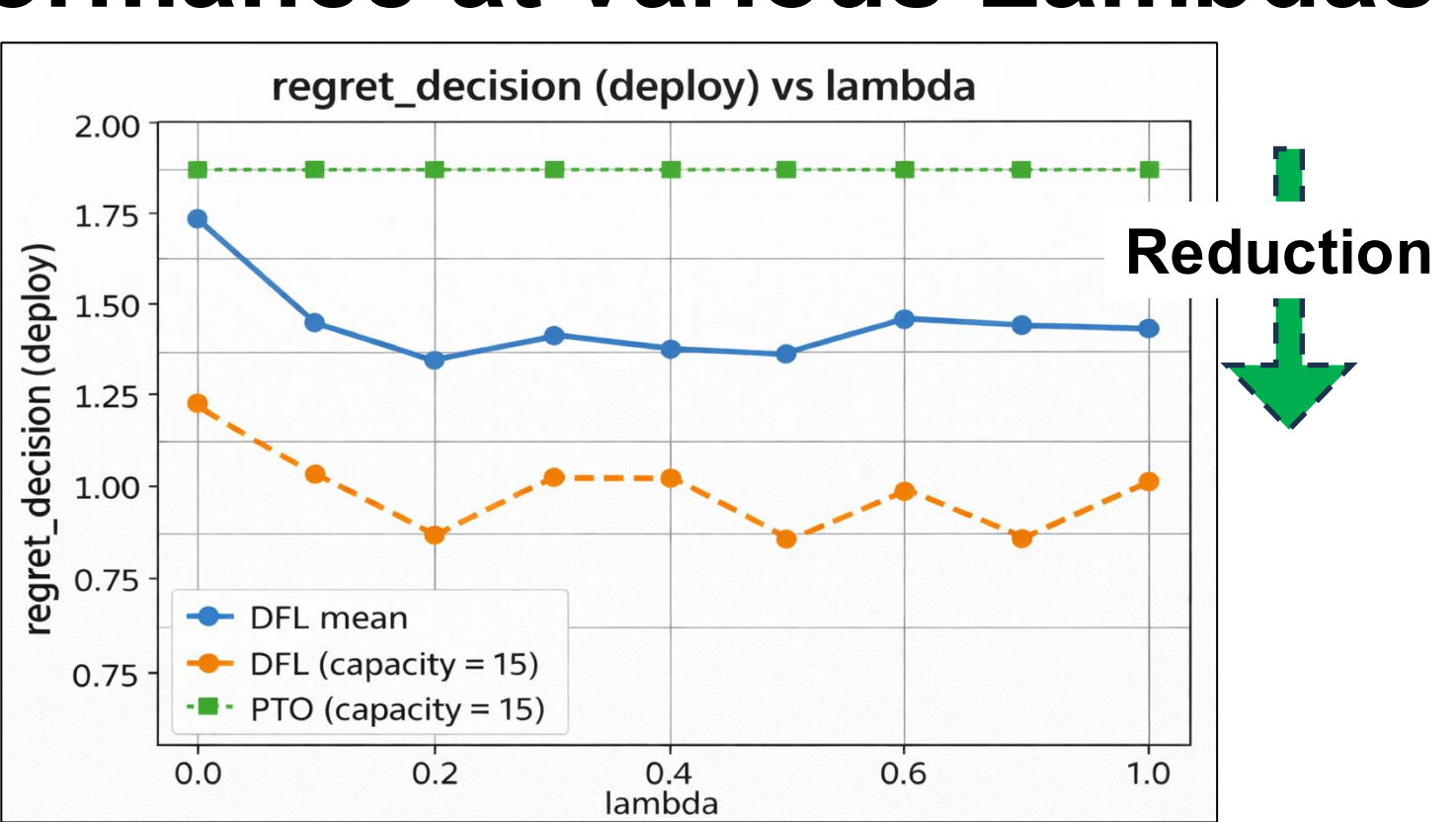
Data Statistics

Knapsack Instance	
Total	50
Label = 1 (illicit)	10
Label = 0	40
Imbalance ratio	0.25

Performance at various Capacities



Performance at various Lambdas



Practical Impact

- Model Applicability:**
- Generalizable to any **business dataset**
 - Metrics are **context-driven**
- Broader Impact:**
- Decision-framework** for investigations
 - Inform **policy reform** and **transparency**

Conclusion & Future Work

- Developed a framework that integrates **prediction & optimization** by aligning model training with downstream knapsack decisions
- The **DFL** outperforms the PTO baseline demonstrating effective resource allocation
- Dynamic Decision:** Model **sequential and budget-adaptive** investigations